

# JOURNAL OF ANIMAL SCIENCE

*The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science*

## **Bayesian analysis of twinning and ovulation rates using a multiple-trait threshold model and Gibbs sampling**

C. P. Van Tassell, L. D. Van Vleck and K. E. Gregory

*J Anim Sci* 1998. 76:2048-2061.

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://jas.fass.org>



**American Society of Animal Science**

[www.asas.org](http://www.asas.org)

# Bayesian Analysis of Twinning and Ovulation Rates Using a Multiple-Trait Threshold Model and Gibbs Sampling<sup>1</sup>

C. P. Van Tassell<sup>\*,2</sup>, L. D. Van Vleck<sup>\*</sup>, and K. E. Gregory<sup>†</sup>

Roman L. Hruska U.S. Meat Animal Research Center, USDA, ARS,

<sup>\*</sup>Lincoln, NE 68583-0908 and Clay Center, NE 68933-0166

**ABSTRACT:** The Multiple-Trait Gibbs Sampler for Animal Models programs were extended to allow analysis of ordered categorical data using a Bayesian threshold model. The algorithm is based on data augmentation, where a value on the unobserved underlying normally distributed variable (liability) is generated in each round of iteration for each categorical observation. The programs allow analysis of several continuous and ordered categorical traits. Categorical traits can have any number of response levels. Models can be different for each trait. The programs were used to analyze twinning and ovulation rates from a herd of cattle selected for twinning rate at the U.S. Meat Animal Research Center. Data included number of calves born at each parturition for the lifetime of a cow and number of eggs ovulated for

several estrous cycles before first breeding as heifers. A total of 6,411 calvings was recorded for 2,087 cows with 83.2% single and 16.8% multiple births. A total of 19,849 ovulations was recorded for 2,332 heifers with 85.2% single and 14.8% multiple ovulations. Mean posterior estimates of heritability and fraction of variance accounted for by permanent environmental effects (PE) were .128 and .103 for twinning rate and .168 and .079 for ovulation rate. Mean posterior estimate of genetic correlation was .808, and correlation of PE effects was .517. Use of a threshold model could allow for more rapid genetic improvement of the twinning herd through improved identification and selection of genetically superior animals because of higher heritability on the underlying scale.

**Key Words:** Threshold Models, Monte Carlo Method, Variance Components, Heritability, Genetic Correlation, Bayesian Theory

©1998 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 1998. 76:2048–2061

## Introduction

Many traits of economic importance in animal breeding are observed as discrete outcomes (e.g., twinning rate, reproductive success, and disease incidence) or subjectively scored in a set of distinct categories (e.g., dystocia and conformation traits). The purpose of this project was to extend the methodology of Sorensen et al. (1995) for univariate threshold models to implement a Gibbs sampling (GS) algorithm for a multiple-trait model for categorical variables that have binomial or multinomial outcomes and for any combination of categorical and

continuous traits using data with any combination of missing traits. A secondary goal was to extend the Multiple-Trait Gibbs Sampler for Animal Models (MTGSAM) (Van Tassell and Van Vleck, 1996) programs based on these theoretical results. The final goal was to apply these procedures to the analysis of ovulation and twinning rates from a long-term selection project for increased twinning rate at the USDA U.S. Meat Animal Research Center.

## Materials and Methods

### Model Background

**Linear Models.** Several authors have examined the use of GS in linear models. Applications in animal breeding include sire models (Wang et al., 1993), animal models (Wang et al., 1994; Van Tassell et al., 1995), and maternal effects models (Jensen et al., 1994; Van Tassell, 1994). In addition, Van Tassell and Van Vleck (1996) derived results for application of GS to a general multiple-trait linear model. A software package was developed to implement these theoretical

<sup>1</sup>The authors wish to acknowledge Cunshan Wang and Dick Quaas for assistance in correcting the derivation of the joint and fully conditional posterior distributions.

<sup>2</sup>To whom correspondence should be addressed. Current address: Animal Improvement Programs Laboratory, USDA-ARS, Building 263, BARC-East, Beltsville, MD 20705-2350; phone: 301-504-9271; fax: 301-504-8092; E-mail: curtvt@aipl.arsusda.gov

Received June 13, 1997.

Accepted January 16, 1998.

results in a flexible set of FORTRAN programs. The results from those derivations will be used as the basis for the further developments in this study.

**Threshold Models.** Ordered categorical traits are often assumed to be controlled by a continuous underlying variable, which is usually called the liability. The observed outcome is then a function of the liability and fixed unknown thresholds; a categorical value is observed if the liability is contained in the region defined by the thresholds for that value.

The liability-threshold concept was first outlined by Wright (1934) for analysis of number of toes in Guinea pigs. Gianola and Foulley (1983) described a Bayesian analysis using threshold models with known thresholds and variance components. Harville and Mee (1984) characterized the likelihood approach to the threshold model problem, including estimation of thresholds and variance components. Foulley et al. (1983) developed Bayesian methodology for analysis of a categorical trait with two continuous traits with the restriction that traits were assumed to be recorded for all three traits for all animals. Janss and Foulley (1993) extended this methodology to allow for missing data for analysis of one categorical and one continuous trait. Finally, Hoeschele et al. (1995) extended the work of Foulley et al. (1983) and Janss and Foulley (1993) to include a single categorical variable that has multiple categories with several continuous variables and any combination of missing traits.

Albert and Chib (1993) described a GS algorithm for threshold models. Sorensen et al. (1995) applied GS to Bayesian analysis of a single-trait threshold model with multiple categories. Wang et al. (1997) developed a GS algorithm that might be applied to an analysis of a multiple-level categorical trait with a continuous trait. The results of these authors are extended here to the completely general case of any combination of continuous and categorical data with missing observations.

### Bayesian Multivariate Threshold Model

Derivation of the threshold model for the multivariate model with linear and threshold traits was based on the model described by Sorensen et al. (1995) and Wang et al. (1997). The mixed linear model notation used by Van Tassell and Van Vleck (1996) will be used to describe the model assumptions and results on the liability scale.

Let  $y$  be the observed random variable for a categorical trait, then  $g$  is the continuous, underlying, normally distributed random variable corresponding to  $y$  (i.e., the liability). For continuous traits, let  $g$  be the observed random variable (i.e.,  $g = y$ ). Assume trait  $i$  has  $\psi_i$  mutually exclusive and exhaustive categories defined by  $\psi_i + 1$  unknown thresholds. It is usual to define  $t_{i,0} = -\infty$  and  $t_{i,\psi_i} = \infty$ . In addition, because the categories are ordered, the intermediate

thresholds must be ordered (i.e.,  $t_{i,1} < t_{i,2} < \dots < t_{i,\psi_i-1}$ ). Then category  $k$  is defined by the region from  $t_{i,k-1}$  to  $t_{i,k}$ .

On the liability scale, residual effects are assumed to be distributed normally. As described by Van Tassell and Van Vleck (1996), residual effects can be partitioned in groups of traits for which non-zero covariances are allowed. Covariances are 0 for residual effects in different groups. Groups are needed when traits are not observed for the same animals (e.g., milk yield and scrotal circumference). Groups are used here only in the context of assigning residual effects to different (co)variance matrices, and these are not related to genetic groups. For this derivation, a record is a set of observations on one or more traits for an animal, and residual effects for these traits may or may not be independent. Although an animal can have multiple records, residual effects for different records are assumed to be independent.

First, the results of Sorensen et al. (1995) for the conditional distribution of the liability to the multivariate setting are extended:

$$\mathbf{g}_j | \mathbf{s}, \mathbf{R}_j^* \sim N(\mathbf{w}_j \mathbf{s}, \mathbf{R}_j^*),$$

where  $\mathbf{g}_j$  are all traits on the observed (underlying) scale for the continuous (categorical) scale for record  $j$ ;  $\mathbf{s}' = [\beta' \mathbf{u}']$ , where  $\beta$  are the fixed effects and  $\mathbf{u}$  are the random effects;  $\mathbf{w}_j$  are rows of  $\mathbf{W} = [\mathbf{X} \mathbf{Z}]$  that correspond to  $\mathbf{g}_j$ , where  $\mathbf{X}$  and  $\mathbf{Z}$  are appropriately dimensioned incidence matrices relating fixed and random effects to the vector of observations; and  $\mathbf{R}_j^*$  is an  $r_j \times r_j$  matrix of (co)variances of residuals for the traits measured on record  $j$ , where  $r_j$  is the number of traits observed for record  $j$ . Assume that the vector of observations on the liability scale is sorted by trait within group within record; then, for a record with all

traits measured  $\mathbf{R}_j^* = \bigoplus_{i=1}^{\rho} \mathbf{R}_i$ , where  $\mathbf{R}_i$  is the  $t_i \times t_i$  covariance matrix among residuals for group  $i$  of the residual effects,  $t_i$  is the number of observed traits in residual group  $i$ ,  $\rho$  is the number of groups of residual effects, and  $\bigoplus$  is the direct sum operator (Searle, 1982). Note that the vector  $\mathbf{g}$  corresponds to the vector of observations for the linear model; the model assumptions for  $\mathbf{g}$  are the same as those for  $\mathbf{y}$  in Van Tassell and Van Vleck (1996).

Conditional on  $\mathbf{s}$ , the vectors of observations or liabilities for records ( $\mathbf{g}_j$ ) are independent. Therefore, the joint density of the data on the liability scale conditional on  $\mathbf{s}$  is simply the product of densities for each record:

$$\begin{aligned} f(\mathbf{g} | \mathbf{s}, \mathbf{R}) &= \prod_{j=1}^q f(\mathbf{g}_j; \mathbf{w}_j \mathbf{s}, \mathbf{R}_j^*) \\ &= \phi(\mathbf{g}; \mathbf{W} \mathbf{s}, \mathbf{R}), \end{aligned}$$

where  $\phi(\mathbf{x}; \mu, \mathbf{V})$  is the probability density function of the (multivariate) normal distribution,  $\mathbf{x} \sim N(\mu, \mathbf{V})$ , as defined in Van Tassell and Van Vleck (1996),  $\mathbf{R} = \bigoplus_{j=1}^q \mathbf{R}_j^*$ , and  $q$  is the number of records.

To facilitate development of fully conditional distributions needed for GS, the conditional distribution of residuals for traits missing in residual blocks with at least one observation is required. Define  $\mathbf{e}_{(j,i)}$  as the vector of residual effects for the traits in residual group  $i$  and record  $j$ . Assume without loss of generality that the vector of residuals can be partitioned as  $\mathbf{e}'_{(j,i)} = [\mathbf{e}'_{(j,i),m} \ \mathbf{e}'_{(j,i),o}]$ , where  $\mathbf{e}_{(j,i),m}$  is the subvector of missing residuals and  $\mathbf{e}_{(j,i),o}$  is the subvector of residuals for observed traits. Further define  $\mathbf{c}_{(j,i),o}$  as the vector of observations or liabilities for observed data, i.e.,  $\mathbf{c}_{(j,i),o} = \mathbf{g}_{(j,i)}$ , and  $\mathbf{c}_{(j,i),m} = \mathbf{e}_{(j,i),m}$ . Then

$$\mathbf{c}' = [\mathbf{g}'_{(1,1)} \ \mathbf{e}'_{(1,1),m} \ \mathbf{g}'_{(1,2)} \ \mathbf{e}'_{(1,2),m} \ \cdots \ \mathbf{e}'_{(q,\rho),m}]$$

and  $\mathbf{e}'_m = [\mathbf{e}'_{(1,1),m} \ \mathbf{e}'_{(1,2),m} \ \cdots \ \mathbf{e}'_{(q,\rho),m}]$ . If all or no traits are observed in a residual group  $i$  for record  $j$ , then  $\mathbf{e}_{(j,i),m}$  is a null vector. That is, missing residuals are only considered if at least one observation is present in a residual group. Similarly,  $\mathbf{g}_{(j,i)}$  is a null vector if no observations are measured for group  $i$  in record  $j$ . These results allow specification of the conditional distribution of  $\mathbf{c}$ :

$$f(\mathbf{c} | \beta, \mathbf{u}, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_\rho) = \phi(\mathbf{c}; \mathbf{W}\mathbf{s}, \mathbf{R})$$

$$\propto |\mathbf{R}|^{-1/2} \times \exp\left\{-\frac{1}{2}(\mathbf{c} - \mathbf{W}\mathbf{s})' \mathbf{R}^{-1}(\mathbf{c} - \mathbf{W}\mathbf{s})\right\},$$

where  $\mathbf{X}$ ,  $\mathbf{Z}$  and  $\mathbf{W} = [\mathbf{X} \ \mathbf{Z}]$  have been redefined to include additional rows of zeros inserted that correspond to missing observations,  $\mathbf{R} = \bigoplus_{j=1}^q \mathbf{R}_j^*$  has been redefined to account for missing observations,

$\mathbf{R}_j^* = \bigoplus_{i \in i_e} \mathbf{R}_i$ , and  $i_e$  includes all residual groups for which record  $j$  has at least one observation.

The next distribution needed is that of the observed values,  $\mathbf{y}$ , conditional on the continuous observations and liabilities,  $\mathbf{c}$ . The relationship between the two variables is unusual because the values of  $\mathbf{y}$  are known if the values of  $\mathbf{c}$  and  $\mathbf{t}$  are known, but the reverse is not true. First, let  $\mathbf{t}' = [\mathbf{t}'_1 \ \mathbf{t}'_2 \ \cdots \ \mathbf{t}'_\tau]$ , where  $\mathbf{t}_i$  is the vector of thresholds for trait  $i$  and  $\tau$  is the total number of traits. For convenience, assume  $\mathbf{t}_i$  is a null length vector for continuous traits. Then if record  $j$  includes an observation for categorical trait  $i$

$$P(y_{ji} = k | g_{ji}, \mathbf{t}) = \begin{cases} 1 & \text{if } t_{i,k-1} < g_{ji} < t_{i,k} \\ 0 & \text{otherwise} \end{cases},$$

and, therefore,

$$f(y_{ji} | g_{ji}, \mathbf{t}) = \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-1} < g_{ji} < t_{i,k}),$$

where  $y_{ji}$  ( $g_{ji}$ ) is the observed (underlying) random variable for trait  $i$  and record  $j$ , and  $I(\cdot)$  is the indicator function. The indicator function has a value of 1 if the evaluated expression is true and a value of 0 otherwise. By definition, for continuous trait  $i$ ,  $f(y_{ji} | g_{ji}, \mathbf{t}) = 1$ . Then, because the observed values are conditionally independent,

$$\begin{aligned} f(\mathbf{y} | \mathbf{g}, \mathbf{t}) &= \prod_{j,i} f(y_{ji} | g_{ji}, \mathbf{t}) \\ &= \prod_{j,i \in i_t} \left[ \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-1} < g_{ji} < t_{i,k}) \right], \end{aligned}$$

where  $i_t$  are categorical traits.

Next, assumptions for the prior distribution are needed. The prior distribution for thresholds assumes that thresholds are distributed as order statistics from a uniform distribution on the interval  $[t_{\min}, t_{\max}]$  (Sorensen et al., 1995; Wang et al., 1997). Specifically,

$$\begin{aligned} f(\mathbf{t}) &= \prod_{i=1}^{\tau} f(\mathbf{t}_i) \\ &= \prod_{i=1}^{\tau} (\psi_i - 1)! \left( \frac{1}{t_{i,\max} - t_{i,\min}} \right)^{\psi_i - 1} I(\mathbf{t}_i \in \mathbf{T}_i), \end{aligned}$$

where

$$\mathbf{T}_i = \left\{ (t_{i,1}, t_{i,2}, \dots, t_{i,\psi_i}) \mid (t_{i,\min} \leq t_{i,1} \leq t_{i,2} \leq \dots \leq t_{i,\psi_i} \leq t_{i,\max}) \right\}.$$

Fixed effects were assumed to have flat prior distributions (i.e.,  $f(\beta) \propto \text{constant}$ ). The random effects were assumed to be normally distributed. Additionally, a known covariance structure corresponding to the numerator relationship matrix was assumed for the genetic effects. Let  $\mathbf{u}' = [\mathbf{u}'_0 \ \mathbf{u}'_1 \ \cdots \ \mathbf{u}'_\gamma]$ , where  $\mathbf{u}_0$  corresponds to the additive genetic effects,  $\mathbf{u}_i$  corresponds to the uncorrelated random effects in group  $i$  for  $i > 0$ , and  $\gamma$  is the number of groups of uncorrelated random effects. As described by Van Tassell and Van Vleck (1996), groups are used only in assignment of uncorrelated random to different (co)variance matrices to force covariances to be zero between specific effects across traits (i.e., these are not genetic groups). Assume  $\mathbf{u}_0$  is ordered by animal within trait and the traits are ordered, and that traits are ordered with direct genetic effects in order for all traits followed by correlated genetic effects for traits that have them. Then  $\mathbf{G}$  is a  $d_0 \times d_0$  matrix that describes the genetic (co)variances among the traits for an animal, where  $d_0$  is the number of traits plus the number of traits with correlated genetic effects. Let  $\mathbf{A}$  be an  $n \times n$  matrix that describes the covariance

structure among genetic effects within traits; typically  $\mathbf{A}$  is the numerator relationship matrix, and  $n$  is the number of animals in the pedigree. The vector  $\mathbf{u}_i$  for  $i > 0$  corresponds to random effects that are uncorrelated with the genetic effects and the other random effects. Then  $\mathbf{D}_i$  is the  $d_i \times d_i$  matrix of (co)variances among the random effects across traits in group  $i$  for an animal, where  $d_i$  is the number of uncorrelated random effects in group  $i$  (i.e., the number of traits represented in group  $i$ ). Let  $n_i$  be the number of levels for each of the  $d_i$  uncorrelated random effects in group  $i$ . The partitioned form of matrix  $\Sigma = \text{Var}(\mathbf{u})$  can be written as

$$\Sigma = \begin{bmatrix} \mathbf{G} \otimes \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \bigoplus_{i=1}^{\gamma} (\mathbf{D}_i \otimes \mathbf{I}_{n_i}) \end{bmatrix},$$

where  $\otimes$  and  $\oplus$  correspond to direct product and direct sum operators, respectively (e.g., Searle, 1982). Let

$\Sigma_0 = \mathbf{G} \otimes \mathbf{A}$  and  $\Sigma_i = \mathbf{D}_i \otimes \mathbf{I}_{n_i}$ , for  $i = 1, \dots, \gamma$ , so that  $\Sigma$  can be also written as  $\Sigma = \bigoplus_{i=0}^{\gamma} \Sigma_i$ .

Inverted Wishart (**IW**) distributions are used as prior distributions for the (co)variance components mainly for computational simplicity. The IW prior distributions for the matrix of (co)variance components are assumed mutually independent. Let  $\mathbf{G}_0$  ( $\nu_g$ ),  $\mathbf{D}_{0i}$  ( $\nu_{d_i}$ ), and  $\mathbf{R}_{0i}$  ( $\nu_{r_i}$ ) be the prior scale (shape or degree of belief) parameters for genetic, group  $i$  of uncorrelated random, and group  $j$  of residual (co)variances, respectively. In addition, let  $\nu_g^* = \nu_g - d_0 - 1$ ,  $\nu_{d_i}^* = \nu_{d_i} - d_i - 1$ , and  $\nu_{r_i}^* = \nu_{r_i} - t_i - 1$ .

In the MTGSAM programs, the user supplies shape parameter and expected value of the prior distribution of the (co)variance matrix, and then the scale matrix is calculated based on those values. For example, for the genetic (co)variance matrix, if the mean value specified by the user is  $\mathbf{G}^*$ , then  $\mathbf{G}_0^{-1} = \nu_g^* \mathbf{G}^*$ .

### Joint Posterior Distributions

From these assumptions and using Bayes theorem, the joint posterior distribution (i.e., the density of the parameters given the data and the prior information) can be formed:

$$\begin{aligned} & f(\mathbf{c}, \mathbf{t}, \beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R} | \mathbf{y}, \nu_g, \mathbf{G}_0, \nu_{d_1}, \mathbf{D}_{0_1}, \dots, \nu_{d_\gamma}, \mathbf{D}_{0_\gamma}, \nu_{r_1}, \mathbf{R}_{0_1}, \dots, \nu_{r_\rho}, \mathbf{R}_{0_\rho}) \\ & \propto f(\mathbf{y} | \mathbf{g}, \mathbf{t}) \times f(\mathbf{c} | \beta, \mathbf{u}, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_\rho) \times f(\beta) \times f(\mathbf{t}) \times f(\mathbf{u} | \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma) \\ & \quad \times f(\mathbf{G} | \nu_g, \mathbf{G}_0) \times \prod_{i=1}^{\gamma} [f(\mathbf{D}_i | \nu_{d_i}, \mathbf{D}_{0_i})] \times \prod_{i=1}^{\rho} [f(\mathbf{R}_i | \nu_{r_i}, \mathbf{R}_{0_i})] \\ & \propto \prod_{j,i \in i_t} \left[ \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-i} < c_{ji} < t_{i,k}) \right] \times |\mathbf{R}|^{-1/2} \times \exp \left\{ -\frac{1}{2} (\mathbf{c} - \mathbf{W}\mathbf{s})' \mathbf{R}^{-1} (\mathbf{c} - \mathbf{W}\mathbf{s}) \right\} \times f(\mathbf{t}) \\ & \quad \times |\mathbf{G}|^{-n/2} \times \exp \left\{ -\frac{1}{2} \mathbf{u}_0' (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{u}_0 \right\} \times \prod_{i=1}^{\gamma} \left[ |\mathbf{D}_i|^{-n_i/2} \times \exp \left\{ -\frac{1}{2} \mathbf{u}_i' (\mathbf{D}_i^{-1} \otimes \mathbf{I}_{n_i}) \mathbf{u}_i \right\} \right] \\ & \quad \times |\mathbf{G}|^{-\frac{1}{2}(\nu_g^* + d_0 + 1)} \times \exp \left\{ \text{tr} \left( -\frac{\nu_g^*}{2} \mathbf{G}_0 \mathbf{G}^{-1} \right) \right\} \times \prod_{i=1}^{\gamma} \left[ |\mathbf{D}_i|^{-\frac{1}{2}(\nu_{d_i}^* + d_i + 1)} \times \exp \left\{ \text{tr} \left( -\frac{\nu_{d_i}^*}{2} \mathbf{D}_{0_i} \mathbf{D}_i^{-1} \right) \right\} \right] \\ & \quad \times \prod_{i=1}^{\rho} \left[ |\mathbf{R}_i|^{-\frac{1}{2}(\nu_{r_i}^* + t_i + 1)} \times \exp \left\{ \text{tr} \left( -\frac{\nu_{r_i}^*}{2} \mathbf{R}_{0_i} \mathbf{R}_i^{-1} \right) \right\} \right] \end{aligned}$$

Then combining terms,

$$\begin{aligned} & f(\mathbf{c}, \mathbf{t}, \beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R} | \mathbf{y}, \nu_g, \mathbf{G}_0, \nu_{d_1}, \mathbf{D}_{0_1}, \dots, \nu_{d_\gamma}, \mathbf{D}_{0_\gamma}, \nu_{r_1}, \mathbf{R}_{0_1}, \dots, \nu_{r_\rho}, \mathbf{R}_{0_\rho}) \\ & \propto f(\mathbf{t}) \times \prod_{j,i \in i_t} \left[ \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-i} < c_{ji} < t_{i,k}) \right] \\ & \quad \times |\mathbf{G}|^{-\frac{1}{2}(n + \nu_g^* + d_0 + 1)} \times \exp \left\{ -\frac{1}{2} (\mathbf{u}_0' (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{u}_0 + \text{tr}(\nu_g^* \mathbf{G}_0 \mathbf{G}^{-1})) \right\} \\ & \quad \times \prod_{i=1}^{\gamma} \left[ |\mathbf{D}_i|^{-\frac{1}{2}(n_i + \nu_{d_i}^* + d_i + 1)} \times \exp \left\{ -\frac{1}{2} (\mathbf{u}_i' (\mathbf{D}_i^{-1} \otimes \mathbf{I}_{n_i}) \mathbf{u}_i + \text{tr}(\nu_{d_i}^* \mathbf{D}_{0_i} \mathbf{D}_i^{-1})) \right\} \right] \\ & \quad \times \prod_{i=1}^{\rho} \left[ |\mathbf{R}_i|^{-\frac{1}{2}(q_i + \nu_{r_i}^* + t_i + 1)} \times \exp \left\{ \text{tr} \left( -\frac{\nu_{r_i}^*}{2} \mathbf{R}_{0_i} \mathbf{R}_i^{-1} \right) \right\} \right] \times \exp \left\{ -\frac{1}{2} (\mathbf{c} - \mathbf{W}\mathbf{s})' \mathbf{R}^{-1} (\mathbf{c} - \mathbf{W}\mathbf{s}) \right\} \end{aligned} \quad [1]$$



Let  $\mathbf{u}'_i = [\mathbf{u}_{i_1} \quad \mathbf{u}_{i_2} \quad \cdots \quad \mathbf{u}_{i_{d_i}}]$  for  $i = 0, \dots, \gamma$ ; then  $\mathbf{S}_0 = \{s_{0,j,k}\}$ , where  $s_{0,j,k} = \mathbf{u}'_{0_j} \mathbf{A}^{-1} \mathbf{u}_{0_k}$ . Similarly,  $\mathbf{S}_i = \{s_{i,j,k}\}$ , where  $s_{i,j,k} = \mathbf{u}'_{i_j} \mathbf{u}_{i_k}$ ,  $\mathbf{Q}_i = \sum_{j=1}^{q_i} \mathbf{e}_{(j,i)} \mathbf{e}'_{(j,i)}$ , and  $q_i$  is the number of records with data observed in residual group  $i$ . Using these new definitions, an alternative form of the joint posterior density is

$$\begin{aligned} & f(\mathbf{c}, \mathbf{t}, \beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R} | \mathbf{y}, \nu_g, \mathbf{G}_0, \nu_{d_1}, \mathbf{D}_{0_1}, \dots, \nu_{d_\gamma}, \mathbf{D}_{0_\gamma}, \nu_{r_1}, \mathbf{R}_{0_1}, \dots, \nu_{r_\rho}, \mathbf{R}_{0_\rho}) \\ & \propto f(\mathbf{t}) \times \prod_{j,i \in i_t} \left[ \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-1} < c_{ji} < t_{i,k}) \right] \times |\mathbf{G}|^{-\frac{1}{2}(n+\nu_g+d_0+1)} \times \exp\left(-\frac{1}{2}\text{tr}((\nu_g^* \mathbf{G}_0 + \mathbf{S}_0) \mathbf{G}^{-1})\right) \\ & \times \prod_{i=1}^{\gamma} \left[ |\mathbf{D}_i|^{-\frac{1}{2}(n_i+\nu_{d_i}+d_i+1)} \times \exp\left(-\frac{1}{2}\text{tr}((\nu_{d_i}^* \mathbf{D}_{0_i} + \mathbf{S}_i) \mathbf{D}_i^{-1})\right) \right] \\ & \times \prod_{i=1}^{\rho} \left[ |\mathbf{R}_i|^{-\frac{1}{2}(q_i+\nu_{r_i}+t_i+1)} \times \exp\left(-\frac{1}{2}\text{tr}((\nu_{r_i}^* \mathbf{R}_{0_i} + \mathbf{Q}_i) \mathbf{R}_i^{-1})\right) \right]. \end{aligned} \quad [2]$$

### Fully Conditional Densities

A set of fully conditional densities is required to implement GS. These densities are required for each scalar element or subvector of elements in the vector of parameters in the model. Each fully conditional density corresponds to the distribution of specific parameter(s) conditional on all other parameters in the model and the data (i.e., the distribution of the specific parameter if the values of the remaining parameters were known). The fully conditional densities can be derived from the joint posterior density (i.e., from [1] or [2]) by ignoring all terms not involving the parameter(s) of interest and then treating the parameters considered to be known as constants and reorganizing terms retained into the density kernel of a known distribution for the parameter of interest. The kernel of a distribution is the part of the density function containing the random variable that remains when all constants are ignored.

**Liabilities.** First, the fully conditional distribution of the underlying liabilities for categorical data and missing residuals is derived from [1]:

$$\begin{aligned} & f(\mathbf{c} | \mathbf{t}, \beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{y}) \\ & \propto \prod_{j,i} \left[ \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-1} < c_{ji} < t_{i,k}) \right] \times \exp\left\{-\frac{1}{2}(\mathbf{c} - \mathbf{Ws})' \mathbf{R}^{-1} (\mathbf{c} - \mathbf{Ws})\right\} \\ & \propto \prod_{j,i} \left[ \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-1} < c_{ji} < t_{i,k}) \right] \times \phi(\mathbf{c}; \mathbf{Ws}, \mathbf{R}), \end{aligned}$$

where the terms in the product  $\left(\prod_{j,i}\right)$  include only those for categorical data. Recall that  $\mathbf{c}$  includes not only liabilities and observations on continuous traits but also missing residuals.

To specify the form of the fully conditional distribution of the liabilities corresponding to observed values, assume, without loss of generality, that the vector of traits can be partitioned as

$$\mathbf{c}_{j.} = \begin{bmatrix} c_{ji} \\ \mathbf{c}_{-ji.} \end{bmatrix},$$

where  $c_{ji} = g_{ji}$  is the liability for the observed trait  $i$  of record  $j$  and  $\mathbf{c}_{-ji.}$  are the remaining values for record  $j$ . Similarly partition  $\mathbf{W}_{j.}$  and  $\mathbf{R}_{j.}^*$  as

$$\mathbf{W}_{j.} = \begin{bmatrix} \mathbf{w}_{ji} \\ \mathbf{W}_{-ji.} \end{bmatrix}, \text{ and}$$

$$\mathbf{R}_j^* = \begin{bmatrix} r_{11} & \mathbf{r}_{12} \\ \mathbf{r}_{21} & \mathbf{R}_{22} \end{bmatrix}.$$

Let  $\mathbf{e}_{j.} = \mathbf{c}_{j.} - \mathbf{W}_{j.}\mathbf{s}$ , and define  $\mathbf{e}_{ji}$  and  $\mathbf{e}_{-ji.}$  similarly. Define  $\mathbf{c}_{-ji}$  as  $\mathbf{c}$  without  $c_{ji}$ , then, using the form of the conditional normal distribution,

$$f(g_{ji} | \mathbf{t}, \beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{c}_{-ji}, \mathbf{y}) \\ \propto \left[ \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-1} < g_{ji} < t_{i,k}) \right] \times \phi(g_{ji}, \mathbf{w}_{ji}\mathbf{s} + \mathbf{r}_{12}\mathbf{R}_{22}^{-1}(\mathbf{c}_{-ji.} - \mathbf{W}_{-ji.}\mathbf{s}), r_{11} - \mathbf{r}_{12}\mathbf{R}_{22}^{-1}\mathbf{r}_{21}) .$$

The distribution corresponds to a truncated normal distribution with truncation points at  $t_{i,k-1}$  and  $t_{i,k}$  if  $y_{ji} = k$  and zero elsewhere; i.e.,

$$g_{ji} | \mathbf{t}, \beta, \mathbf{u}, \mathbf{R}, \mathbf{c}_{-ji}, \mathbf{y} \sim \text{TN}(\mathbf{w}_{ji}\mathbf{s} + \mathbf{r}_{12}\mathbf{R}_{22}^{-1}(\mathbf{c}_{-ji.} - \mathbf{W}_{-ji.}\mathbf{s}), r_{11} - \mathbf{r}_{12}\mathbf{R}_{22}^{-1}\mathbf{r}_{21}) \\ \text{for } t_{i,k-1} < g_{ji} < t_{i,k} \text{ and } y_{ji} = k,$$

where TN indicates the truncated normal distribution. This distribution can be stated in terms of the residual  $e_{ji} = c_{ji} - \mathbf{w}_{ji}\mathbf{s}$  because there is a one-to-one relationship with the liability given the fixed and random effects:

$$e_{ji} | \mathbf{t}, \beta, \mathbf{u}, \mathbf{R}, \mathbf{c}_{-ji}, \mathbf{y} \sim \text{TN}(\mathbf{r}_{12}\mathbf{R}_{22}^{-1}(\mathbf{c}_{-ji.} - \mathbf{W}_{-ji.}\mathbf{s}), r_{11} - \mathbf{r}_{12}\mathbf{R}_{22}^{-1}\mathbf{r}_{21}) \\ \text{for } t_{i,k-1} - \mathbf{w}_{ji}\mathbf{s} < e_{ji} < t_{i,k} - \mathbf{w}_{ji}\mathbf{s} \text{ and } y_{ji} = k. \quad [3]$$

*Thresholds.* The fully conditional distribution of the thresholds are derived from [1]:

$$f(\mathbf{t} | \beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{c}, \mathbf{y}) \propto \prod_{j, i \in i_t} \left[ \left( \sum_{k=1}^{\psi_i} I(y_{ji} = k) \times I(t_{i,k-1} < g_{ji} < t_{i,k}) \right) \times I(\mathbf{t}_i \in \mathbf{T}_i) \right], \quad [4]$$

where the terms in the product again include only those for categorical traits. From [4], the upper limit for  $t_{i,k}$  is  $\min(g_{ji} | y_{ji} = k+1)$  and the lower limit is  $\max(g_{ji} | y_{ji} = k)$ . Because the density is proportional to a constant, which is the kernel of a uniform distribution, the fully conditional distribution for a threshold is uniform with limits of  $\max(g_{ji} | y_{ji} = k)$  and  $\min(g_{ji} | y_{ji} = k+1)$ , i.e.,

$$f(t_{i,k} | \mathbf{t}_{-ik}, \beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{c}, \mathbf{y}) \propto \frac{1}{\min(g_{ji} | y_{ji} = k + 1) - \max(g_{ji} | y_{ji} = k)},$$

and

$$t_{i,k} | \mathbf{t}_{-ik}, \mathbf{c}, \mathbf{y} \sim U(\max(g_{ji} | y_{ji} = k), \min(g_{ji} | y_{ji} = k+1)). \quad [5]$$

The notation  $U(a,b)$  indicates that a variable has a continuous uniform distribution with minimum  $a$  and maximum  $b$ . For these derivations to hold, data are assumed to be observed for all categories of interest. For example, if calving ease is scored on discrete values from 1 to 5, then it is assumed that there are observations for all levels or that only thresholds between observed categories are of interest. The fully conditional distribution in the general case where empty categories are allowed is given by Albert and Chib (1993) and Sorensen et al. (1995).

*Remaining Effects.* The remaining fully conditional distributions are analogous to those given in Van Tassell and Van Vleck (1996) with  $\mathbf{c}$  replacing  $\mathbf{y}$  (i.e., the vector of liabilities and continuous data replaces the vector of continuous data). Specifically, the fully conditional distribution of the vector of fixed and random effects is

$$\mathbf{s} | \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{c} \sim N(\tilde{\mathbf{s}}, \mathbf{C}^{-1}),$$

$$\text{where } \tilde{\mathbf{s}} = \mathbf{C}^{-1}\mathbf{r}, \quad \mathbf{r} = \mathbf{W}'\mathbf{R}^{-1}\mathbf{c}, \quad \text{and } \mathbf{C} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \Sigma^{-1} \end{bmatrix}.$$

Because of the definitions of  $\mathbf{R}$ ,  $\mathbf{X}$ , and  $\mathbf{Z}$ ,  $\mathbf{C}$  is **not** the coefficient matrix for Henderson's mixed model equations with missing observations, and  $\mathbf{r}$  is not the MME right-hand side because the missing residuals are included in  $\mathbf{c}$ . Then the fully conditional distribution of a scalar or subvector element  $\mathbf{s}_1$  of  $\mathbf{s}$ , where  $\mathbf{s}' = [\mathbf{s}'_1 \quad \mathbf{s}'_2]$ , is

$$\mathbf{s}_1 | \mathbf{s}_2, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{C} \sim N(\mathbf{C}_{11}^{-1} (\mathbf{r}_1 - \mathbf{C}_{12}\mathbf{s}_2), \mathbf{C}_{11}^{-1}), \quad [6]$$

$$\text{where } \mathbf{r}' = [\mathbf{r}'_1 \quad \mathbf{r}'_2] \quad \text{and } \mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}.$$

The blocking algorithm of Van Tassell and Van Vleck (1996) is retained. The fully conditional distribution of a subvector or scalar can be written as

$$\mathbf{s}_i | \mathbf{s}_{-i}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{c} \sim N(\tilde{\mathbf{s}}_i, (\mathbf{P}_i \mathbf{C} \mathbf{P}_i')^{-1}), \quad [7]$$

$$\text{where } \tilde{\mathbf{s}}_i = (\mathbf{P}_i \mathbf{C} \mathbf{P}_i')^{-1} (\mathbf{P}_i \mathbf{r} - \mathbf{P}_i \mathbf{C} \mathbf{P}_{-i}' \mathbf{s}_{-i}), \quad \mathbf{P} \text{ is a permutation matrix such that } \mathbf{P}\mathbf{s} = \begin{bmatrix} \mathbf{P}_i \mathbf{s} \\ \mathbf{P}_{-i} \mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_i \\ \mathbf{s}_{-i} \end{bmatrix},$$

$\mathbf{s}_i$  are the random effects in the block, and  $\mathbf{s}_{-i}$  are the remaining fixed and random effects. Although this form appears complex, it is simply a precise representation of [6], where the permutation of elements is shown here explicitly. The (co)variance matrix is comprised of the appropriate elements of the coefficient matrix corresponding to the rows and columns for the elements in the block, and the mean is equal to the update in a block Gauss-Seidel iteration algorithm.

Fully conditional distributions for missing residuals can be obtained from [1]:

$$f(\mathbf{e}_m | \mathbf{g}, \beta, \mathbf{u}, \mathbf{R}) \propto \exp\left\{-\frac{1}{2} (\mathbf{c} - \mathbf{W}\mathbf{s})' \mathbf{R}^{-1} (\mathbf{c} - \mathbf{W}\mathbf{s})\right\}.$$

Recall that  $\mathbf{e}_{(j,i)}$  includes  $\mathbf{e}_{(j,i),m}$  and  $\mathbf{e}_{(j,i),o}$ . Then for residuals corresponding to observed data,  $\mathbf{e}_{(j,i),o} = \mathbf{g}_{(j,i),o} - \mathbf{w}_{(j,i),o}\mathbf{s}$ . Residual effects for group  $i$  are distributed normally:  $\mathbf{e}_{j,i} | \mathbf{R}_i \sim N(\mathbf{0}, \mathbf{R}_i)$ . Let

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{mm} & \mathbf{R}_{mo} \\ \mathbf{R}_{om} & \mathbf{R}_{oo} \end{bmatrix} \text{ be partitioned as } \mathbf{e}_{ji}, \text{ and then}$$

$$\mathbf{e}_{(j,i),m} | \mathbf{e}_{(j,i),o}, \mathbf{s}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{g} \sim N(\mathbf{R}_{mo} \mathbf{R}_{oo}^{-1} \mathbf{e}_o, \mathbf{R}_{mm} - \mathbf{R}_{mo} \mathbf{R}_{oo}^{-1} \mathbf{R}_{om}). \quad [8]$$

Finally, as described by Van Tassell and Van Vleck (1996), the fully conditional distributions of the (co)variance matrices can be derived using form [2] of the joint posterior distribution. The fully conditional distributions for the genetic (co)variance matrix, each group of uncorrelated random effects, and each group of residual effects are

$$\mathbf{G} | \mathbf{s}, \mathbf{G}_0, \nu_g, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{c} \sim IW\left(\left(\nu_g^* \mathbf{G}_0 + \mathbf{S}_0\right)^{-1}, n + \nu_g\right), \quad [9]$$

$$\mathbf{D}_i | \mathbf{s}, \mathbf{D}_{0_i}, \nu_{d_i}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_{i-1}, \mathbf{D}_{i+1}, \dots, \mathbf{D}_\gamma, \mathbf{R}, \mathbf{c} \sim IW\left(\left(\nu_{d_i}^* \mathbf{D}_{0_i} + \mathbf{S}_i\right)^{-1}, n_i + \nu_{d_i}\right), \text{ and} \quad [10]$$

$$\mathbf{R}_i | \mathbf{e}_i, \mathbf{R}_{0_i}, \nu_{r_i}, \mathbf{s}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_\gamma, \mathbf{R}_1, \dots, \mathbf{R}_{i-1}, \mathbf{R}_{i+1}, \dots, \mathbf{R}_\rho, \mathbf{c} \sim IW\left(\left(\nu_{r_i}^* \mathbf{R}_{0_i} + \mathbf{Q}_i\right)^{-1}, q_i + \nu_{r_i}\right), \quad [11]$$

respectively.



### Starting Values

For the continuous traits, starting (co)variances are supplied by the user for those components. The starting values correspond to the means of prior distributions if using informative prior distributions for the (co)variances. Gauss-Seidel iteration using the starting variance components is used to calculate starting values for fixed and random effects. The user specifies maximum number of rounds of iteration and convergence criterion.

For categorical traits, starting values for fixed and random effects and thresholds pose a challenge for the MTGSAM programs. Ideally, the starting values should be solutions to a threshold model. Hoeschele and Tier (1995), for example, suggest using maximum *a priori* (**MAP**) estimates evaluated at the approximate marginal maximum likelihood (**MML**) estimates of the (co)variance components. This approach is not practical in development of these programs, because general MML and MAP estimates would dramatically increase computing demands and algorithm complexity for a general model. An approximation based on a linear model was chosen. A score based on the cumulative distribution function (**CDF**) of the normal distribution suggested by Djemali et al. (1987) was used. The score is calculated as the median of the CDF for a category. If the cumulative fractions of scores for categories 1 to  $\psi_i$  for trait  $i$  are  $p_{i1}, p_{i2}, \dots, p_{i\psi_i}$  then  $y_{ji}^* = \Phi^{-1}\left(\frac{1}{2}(p_{i,k} - p_{i,k-1})\right)$ , where  $y_{ji}^*$  is the normalized score,  $\Phi^{-1}(\cdot)$  is the standard inverse cumulative normal function, and  $p_{i0}$  is 0. Rank correlation of solutions using the normalized score with solutions using a threshold model was .99 in the study by Djemali et al. (1987). Solutions were calculated using Gauss-Seidel iteration with the normalized scores and the starting variance components. The starting values for the thresholds are based on the approximate liabilities using the linear model. These approximate liabilities are calculated as the sum of the solutions for the fixed and random effects for an observation, i.e.,  $c_{ji}^* = w_{ji}s$ . The starting threshold is the midpoint of the extremes of the adjacent categories:

$$t_{ik}^* = \frac{1}{2}(\min(c_{ji}^* | y_{ji} = k + 1) + \max(c_{ji}^* | y_{ji} = k)).$$

### Constraints

Because the underlying scale is arbitrary for categorical variables with respect to mean and variance, constraints must be imposed to guarantee identifiability. Residual variance of 1 and  $t_{1,i} = 0$  are usually assumed. Sorensen et al. (1995) refer to this as the “standard” parameterization. As described by Sorensen et al. (1995), an alternate parameterization

can be used with multiple categories: residual variance is assumed unknown (and estimated),  $t_{1,i} = 0$ , and  $t_{2,i} = 1$ . Sorensen et al. (1995) found that the alternate parameterization may allow for increased rate of convergence when a Markov Chain Monte Carlo algorithm (e.g., GS) is used to characterize the posterior distributions. For binary data, the standard parameterization is used because only one threshold exists. For categorical traits with at least three categories, both parameterizations are possible.

### Implementation of the Gibbs Sampler

Using the fully conditional distributions, [3] to [11], the GS algorithm used can be outlined as follows:

1. Input or calculate starting values for all variables as described.
2. Generate liabilities for categorical observations from [3].
3. Generate thresholds from [4].
4. Generate fixed effects from [6].
5. Generate genetic effects and blocked uncorrelated random effects from [7].
6. Generate uncorrelated random effects not in a block from [6].
7. Calculate residual effects for traits with observations, and generate missing residuals from [8].
8. Calculate quadratics for genetic effects,  $S_0$ , from  $u_0' A^{-1} u_0$ .
9. Generate  $G$  from [9].
10. Calculate quadratics for each group of uncorrelated random effects,  $S_i$ , from  $u_{ij}' u_{ik}$ .
11. Generate each  $D_i$  from [10].
12. Calculate quadratics for each group of residual effects,  $Q_i$ , from  $e_{ij}' e_{ij}$ .
13. Generate each  $R_i$  from [11] and rescale them as necessary.
14. Repeat steps 2 through 13 many times.

### MTGSAM Flexibility

The MTGSAM programs were designed to be as general as possible. The programs allow analysis of several continuous and categorical variables. Categorical variables can have any number of levels. The categorical levels must be ordered but are not required to be sequential or to start at 1. The program allows any number of fixed and random effects for categorical or continuous traits. The model can be different for each trait.

The MTGSAM programs allow for use of standard or alternate parameterization of the thresholds and residual variances. Standard parameterization is assumed for binary data. For categorical traits with at least three categories the MTGSAM user can choose from the two possible parameterizations. When using

the standard parameterization sampling the residual variance is done as usual, but after sampling the residual variances are set to 1 and the covariances are scaled so that correlations are identical to those sampled. Although the scaling procedure is *ad hoc* in nature, the sampling sequence should result in a Markov Chain because of the arbitrary underlying scale for categorical traits.

Finally, the MTGSAM programs can be used for sire models by considering the genetic effect as a sire effect, and a sire-maternal grandsire model can be fit by considering maternal grandsire effect as the correlated genetic effect. For sire-maternal grandsire models that are fit using this method, the existence of maternal effects is assumed implicitly because sire and maternal grandsire effects are fit separately. The use of sire or sire-maternal grandsire models may be especially useful for categorical data with low frequencies, because the Gibbs sampler may not converge in these situations using animal models (Hoeschele and Tier, 1995).

### *Analysis of Twinning and Ovulation Rate Data*

**Data.** The programs were used to analyze twinning and ovulation rates from a herd of cattle selected for twinning rate at the U.S. Meat Animal Research Center (**USMARC**). A brief summary of the experiment follows, but previous papers from the twinning and ovulation rate project have completely described the foundation animals and general management of the animals in the project (Echternkamp et al., 1990) as well as the collection of ovulation rate information (Gregory et al., 1990). The selection experiment was initiated in 1981 at USMARC to increase twinning rate of beef cattle in an effort to improve the economic efficiency of production. Foundation cows were obtained from two sources: 96 cows with records of two or more twin calvings were acquired from industry, and 211 females with records of twinning in other projects were transferred to the twinning project. Foundation sires were obtained from several sources: semen from three Swedish Friesian and two Norwegian Red bulls was obtained, semen from two USMARC bulls (a Pinzgauer and a Charolais) with daughters that had produced 8 to 10% twins during other projects, and randomly chosen unproven bulls out of 36 foundation cows. Semen from three additional Swedish Friesian sires was introduced in 1988. Multiple ovulation and embryo transfer were used to augment the first generation. Breeds with significant representation included Holstein, Swedish Friesian, Simmental, Pinzgauer, Charolais, Swedish Red and White, Norwegian Red, Hereford, and Angus. About 750 females calved each year during two calving seasons (spring and fall). Approximately the top 25% of females for predicted breeding value (**PBV**) for twinning were mated to progeny-tested, proven sires. The remaining females were mated to young, un-

proven sires. Approximately 30 males were identified by PBV each year for progeny testing through a two-stage selection procedure.

Initial study of evaluation of ovulation rate as a correlated trait began with females born in fall 1985. At that time, approximately half of the daughters of each bull were randomly chosen to be palpated per rectum for 8 to 10 estrous cycles and bred to calve at approximately 2.5 yr of age. The remaining female contemporaries were bred to calve at approximately 2 yr of age. Preliminary results supported determination of ovulation rate for all heifers beginning with those born in fall 1986.

Van Vleck and Gregory (1996) presented a REML analysis of a data set using a linear model that included nearly all of the data analyzed in this study. An additional season of ovulation data was added to their data. A more recent analysis of similar data with early data excluded was presented by Gregory et al. (1997).

Data included number of calves born at each parturition for the lifetime of a cow and number of eggs ovulated for several estrous cycles before first breeding as heifers. Observations of triplets born ( $n = 25$ ) or three eggs ovulated ( $n = 29$ ) were included as counts of two (i.e., both traits were considered binary). A total of 6,411 calvings was recorded for 2,087 cows with 83.2% single and 16.8% multiple births. A total of 19,849 ovulations was recorded for 2,332 heifers with 85.2% single and 14.8% multiple ovulations. Approximately 24, 20, 19, 15, 11, and 11% of the cows with twinning data had 1, 2, 3, 4, 5, and  $> 5$  observed calvings, respectively, and 10, 7, 23, 34, 16, and 10% of the heifers with ovulation data had  $< 7$ , 7, 8, 9, 10, and  $> 10$  recorded estrous cycles, respectively.

**Model.** The bivariate animal models used to analyze these data correspond to those used by Van Vleck and Gregory (1996). The model for twinning rate included a fixed factor for effects of combinations of year, season, and age at parturition and random factors for animal (additive genetic) and permanent environmental (**PE**) effects of the cow. The model for ovulation rate included fixed effects of birth year-season, age at observation ( $< 12$  mo, 12 to 13 mo, 14 to 15 mo, 16 to 17 mo, and  $> 17$  mo), and calendar month at observation and random factors for animal and PE effects of the heifer. The residual correlation was assumed to be 0 because calving and ovulation data were not recorded simultaneously.

Genetic groups were included to account for breed differences and selection periods. Westell (1984) groups were assigned based on subpopulations represented in the base population. The Westell group effects were included in the relationship matrix using the algorithm described by Westell et al. (1988).

**Gibbs Sampler and Posterior Estimation.** Two independent chains were computed, and samples from

the two chains were combined for analysis. Each chain was run for a total of 225,000 rounds with the first 25,000 rounds discarded as the burn-in or initialization period. The burn-in period was based on subjective evaluation of plots of values from the Gibbs chain. Although the chains seemed to have converged by 5,000 rounds of iteration, a conservative burn-in period of 25,000 rounds was used. Flat prior distributions were used for variance components. Heritability of .185 and fraction of PE variance of .074 were assumed for starting values for both traits. Starting values of .80 and .50 were used for genetic and PE correlations, respectively.

Lag correlations of values sampled from the Gibbs chain were calculated for a range of intervals between sampled values. Lag correlations were fit to a nonlinear curve that assumed lag correlation decreased geometrically with interval between samples; the equation fit  $\rho_i = \rho_1^i$ , where  $\rho_i$  is the correlation of samples drawn  $i$  rounds apart for a given parameter. The curve was fit (i.e.,  $\rho_1$  was estimated) using PROC NLIN of SAS (1989). After estimates of  $\rho_1$  were obtained, thinning intervals were calculated to obtain lag correlations of a specific level. If a lag correlation of  $\alpha$  ( $0 < \alpha < 1$ ) is desired, and a geometric decrease in correlations is assumed, the thinning interval ( $n_\alpha$ ) required to obtain that correlation of thinned samples can be written as  $\alpha = (\rho_1)^{n_\alpha}$ . Solving for  $n_\alpha$ ,

$$n_\alpha = \log(\alpha)/\log(\rho_1).$$

The advantage of this approach is that the cutoff points for a specific lag correlation are easily determined.

Because of space limitations, samples were taken at 10-round intervals after the burn-in period. Posterior distribution estimates were calculated using the unthinned samples from the Gibbs chain. Although these samples have a high autocorrelation, this provides an unbiased estimate of the posterior distribution. The posterior distributions were estimated using the general average shifted histogram algorithm of Scott (1992). Posterior distribution estimates were also used to estimate modes and credible sets (Bayesian confidence intervals).

Other researchers have observed bias in heritability estimates when using GS and threshold models (Hoeschele and Tier, 1995; Moreno et al., 1997). Heritability tends to be biased upward when the amount of information per fixed effect is small. However, models used for these data included relatively small numbers of fixed effects: twinning rate included one fixed effect with 96 levels, and ovulation rate included three cross-classified fixed effects with 5, 12, and 17 levels. Because fixed effects should be relatively well estimated, estimates should have little bias.

Point estimates of means for each parameter were calculated as the means of samples or functions of samples taken at 10-round intervals. Variability of point estimates due to the length of the Gibbs chain (i.e., Monte Carlo error) were estimated. Monte Carlo standard deviation was estimated as the standard error of the mean of samples that were thinned to a lag correlation of .01 (i.e., nearly independent). The thinning rate was based on the nonlinear estimation of lag correlation.

**REML Linear Model Analyses.** The analyses of Van Vleck and Gregory (1996) were repeated for the data because an additional season of ovulation data was added after their study was completed. The models used were the same as in that study, and parameter estimates were obtained using the Multiple-Trait Derivative-Free REML (**MTDFREML**) programs (Boldman et al., 1995). The effects included in the models were the same as those in the threshold models. The programs were restarted and allowed to iterate until a high level of convergence was attained. Convergence was declared when the variance of the simplex values ( $-2 \log$  likelihoods) was  $< 10^{-10}$ . To ensure global convergence, the programs were restarted with previous converged solutions until convergence occurred at the same maximum.

Heritability estimates from the linear model analysis were rescaled to the underlying scale using the transformation suggested by Robertson in Dempster and Lerner (1950). Let  $p_i$  be the frequency of outcome 1 for trait  $i$  (e.g., twinning rate),  $h_{iOBS}^2$  be heritability on the observed (binary) scale, and  $h_{iCONT}^2$  be heritability on the underlying continuous scale. Then

$$h_{iCONT}^2 = \frac{h_{iOBS}^2 \times p_i (1 - p_i)}{[\phi(\Phi^{-1}(p_i); 0, 1)]^2}.$$

The same transformation was also applied to the fraction of variance accounted for by PE.

## Results and Discussion

Point estimates for heritabilities, fractions of variance, and correlations are presented in Table 1 for the Bayesian posterior distributions and for the linear model (**LM**) REML analysis. Posterior density estimates are shown in Figures 1 and 2. Table 1 also contains 95% credible sets for posterior means.

Mean posterior heritability estimates were .128 and .168 for twinning and ovulation rate. The transformed heritabilities from the LM analysis were .098 and .175. The transformed LM heritability estimate was similar to the threshold estimate for twinning rate. For ovulation rate, however, the transformed LM estimates was only 77% of the threshold estimate. The

Table 1. Marginal posterior means and lower and upper cutoffs for 95% credible sets from Bayesian threshold model, REML estimates from linear models, and REML estimates rescaled to underlying scale for heritabilities ( $h^2$ ), fractions of variance accounted for by permanent environmental ( $c^2$ ) and residual ( $r^2$ ) effects for twinning (T) and ovulation (O), and genetic ( $r_g$ ) and permanent environmental ( $r_c$ ) correlations of effects across traits

Parameter	Threshold model			Linear model	Linear rescaled
	Mean	Lower cutoff	Upper cutoff		
$h_T^2$	.128	.085	.173	.044	.098
$h_O^2$	.168	.124	.215	.074	.175
$r_g$	.808	.690	.920	.709	—
$c_T^2$	.103	.050	.154	.029	.065
$c_O^2$	.079	.042	.115	.041	.097
$r_c$	.517	.183	.859	.773	—
$r_T^2$	.769	.721	.815	.926	—
$r_O^2$	.753	.726	.780	.885	—

heritabilities indicate that significant genetic variation exists for both traits. Estimates of fraction of variation associated with PE were smaller than heritabilities but accounted for substantial variation for both traits. Mean estimates of fractions of variance differed from modal estimates by  $<.005$  except for PE fraction (mean [mode] estimates of .103 [.098]), which indicates that the posterior means were relatively symmetric (Figure 1).

Mean posterior estimates of genetic correlation were .808 for the threshold model, and the LM estimate was .709. The posterior mode estimate was .835. The difference between mean and modal estimates reflects some lack of symmetry in posterior distribution estimates (Figure 2). Although the estimates from the linear and threshold models differed, the 95% credible set from the threshold model included the LM estimates. The high genetic correlation between the traits indicates that ovulation rate is closely related to twinning rate and can be used successfully for selection before twinning rate can be observed.

Mean posterior estimates of PE correlation was .517, and the LM estimate was .773. Estimates of posterior mode were .538 and again reflect asymmetry (Figure 2). The credible set is quite large for the estimate of PE correlation, which indicates poor precision for this parameter. The wide credible set results, in part, because of data structure. Although the data set included many repeated records for both traits (as confidence intervals for PE fraction indicate),  $<25\%$  of the animals have multiple observations for both calving and ovulation rate.

Estimates of Monte Carlo standard deviations (MCS D) are presented in Table 2. The MCS D were  $<.005$  for all parameters except correlations, which were  $<.01$  for genetic correlations and  $<.05$  for PE correlation. Two factors resulted in larger MCS D for correlations compared with variance fractions. First,

the number of independent samples was smaller than for variance fractions because of higher correlations among sequential samples (Table 2); therefore, variance fractions had higher thinning intervals. Second, the variance of the posterior distributions was larger for the correlations. The differences in variance can be seen in Figures 1 and 2 by comparing the range of values for the x-axis on the graphs.

### Program Availability

The MTGSAM programs are available through the Internet or from the authors. The MTGSAM home page is available via the Animal Geneticists Discus-

Table 2. Monte Carlo standard deviations (MCS D), number of samples used to calculate MCS D (N), and lag correlations with 500 rounds between values sampled from the Gibbs chain (R500) for heritabilities ( $h^2$ ), fractions of variance accounted for by permanent environmental ( $c^2$ ) and residual ( $r^2$ ) effects for twinning (T) and ovulation (O), and genetic ( $r_g$ ) and permanent environmental ( $r_c$ ) correlations of effects across traits

Parameter	Model with genetic groups		
	MCS D	N	R500
$h_T^2$	.0038	56	.054
$h_O^2$	.0026	66	.033
$r_g$	.0085	34	.183
$c_T^2$	.0046	42	.118
$c_O^2$	.0025	54	.063
$r_c$	.0319	32	.198
$r_T^2$	.0036	55	.059
$r_O^2$	.0008	305	.000



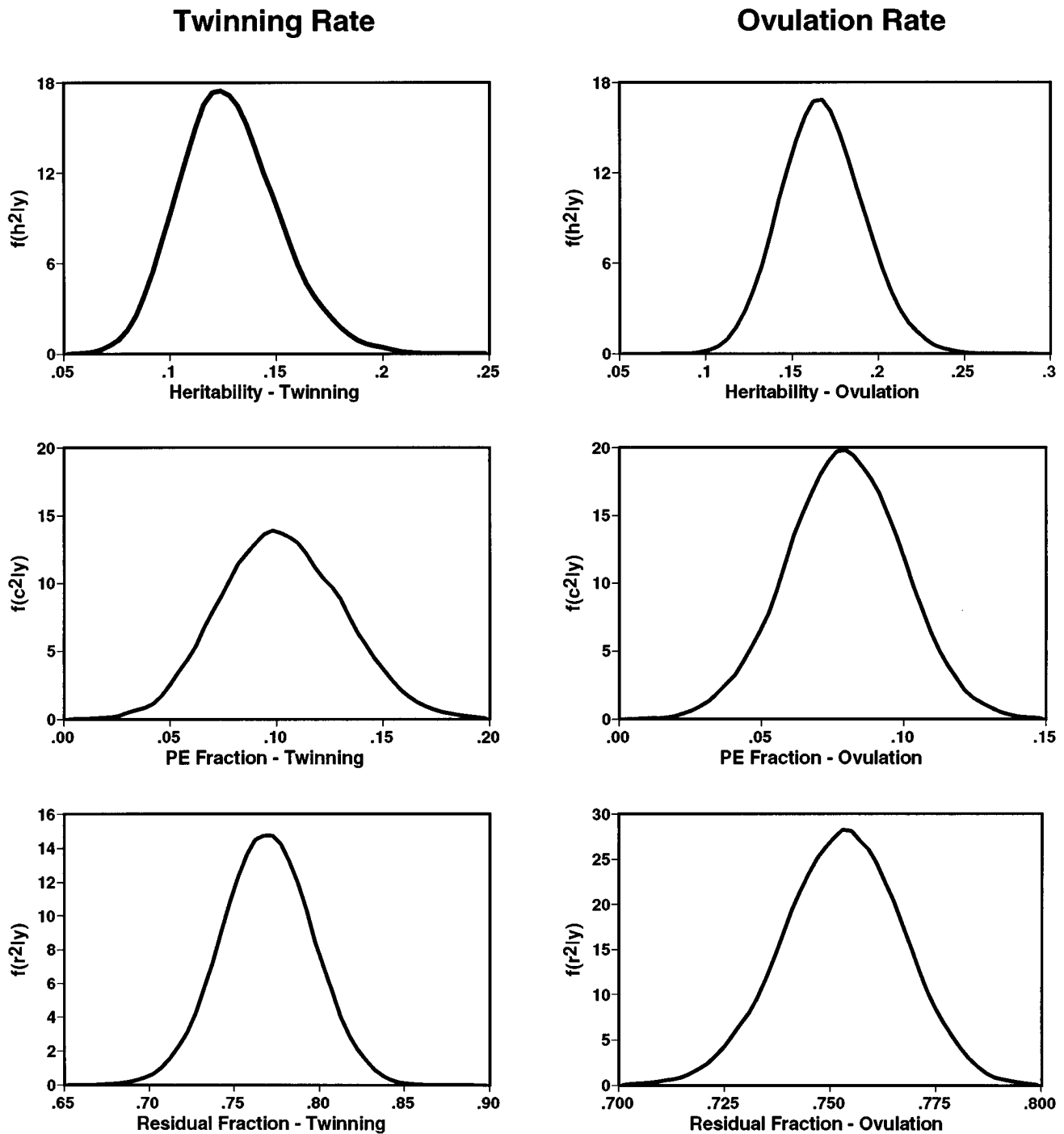


Figure 1. Marginal posterior distributions of fractions of variance for twinning and ovulation rates for genetic, permanent environmental (PE), and residual effects.

sion Group home page, which can be accessed from links on the American Society of Animal Science home page (<http://www.asas.uiuc.edu>).

The MTGSAM programs are intended to be used by those familiar with the concepts of GS, including evaluation of the burn-in period, thinning interval, and convergence. The programs simplify much of the work required to use GS, but users must still make certain that results obtained are sensible.

### Implications

The MTGSAM programs were extended to allow ordered categorical data with a Bayesian threshold model. The improved programs allow simultaneous analysis of continuous and categorical data. The programs estimate (co)variance components, variance fractions, correlations, fixed and random effects, linear functions of effects, and thresholds. Because of use of a

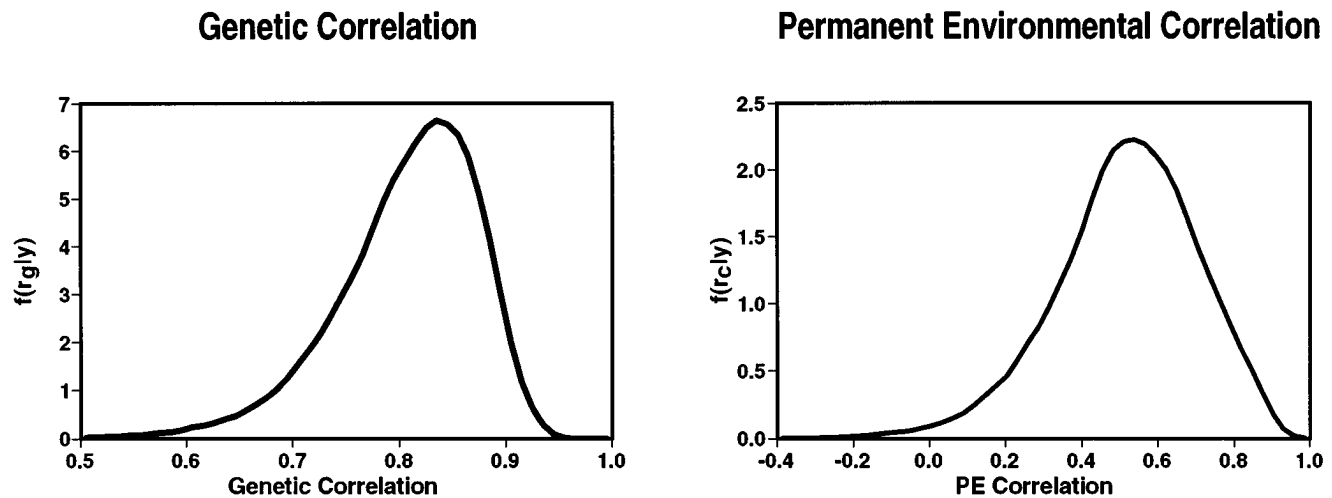


Figure 2. Marginal posterior distributions of genetic and permanent environmental (PE) correlations of twinning and ovulation rates.

more correct model and higher heritability on the underlying scale, use of a threshold model could allow for more rapid genetic improvement of the twinning herd through improved identification and selection of genetically superior animals.

### Literature Cited

- Albert, J. H., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* 88:669–679.
- Boldman, K. G., L. A. Kriese, L. D. Van Vleck, C. P. Van Tassell, and S. D. Kachman. 1995. A Manual for Use of MTDFREML: A Set of Programs To Obtain Estimates of Variances and Covariances [DRAFT]. USDA, ARS, Clay Center, NE.
- Dempster, E. R., and I. M. Lerner. 1950. Heritability of threshold characters. With an appendix by A. Robertson. *Genetics* 35: 212–236.
- Djemali, M., P. J. Berger, and A. E. Freeman. 1987. Ordered categorical sire evaluation for dystocia in Holsteins. *J. Dairy Sci.* 20:2374–2384.
- Echternkamp, S. E., K. E. Gregory, G. E. Dickerson, L. V. Cundiff, R. M. Koch, and L. D. Van Vleck. 1990. Twinning in cattle: II. Genetic and environmental effects on ovulation rate in pubertal heifers and postpartum cows and the effects of ovulation rate on embryonic survival. *J. Anim. Sci.* 68:1877–1888.
- Foulley, J. L., D. Gianola, and R. Thompson. 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genet. Sel. Evol.* 15:401–424.
- Gianola, D., and J. L. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15: 201–224.
- Gregory, K. E., G. L. Bennett, L. D. Van Vleck, S. E. Echternkamp, and L. V. Cundiff. 1997. Genetic and environmental parameters for ovulation rate, twinning rate, and weight traits in a cattle population selected for twinning. *J. Anim. Sci.* 75: 1213–1222.
- Gregory, K. E., S. E. Echternkamp, G. E. Dickerson, L. V. Cundiff, R. M. Koch, and L. D. Van Vleck. 1990. Twinning in cattle: I. Foundation animals and genetic and environmental effects on twinning rate. *J. Anim. Sci.* 68:1867–1876.
- Harville, D. A., and R. W. Mee. 1984. A mixed-model procedure for analyzing ordered categorical data. *Biometrics* 40:393–408.
- Hoeschele, I., and B. Tier. 1995. Estimation of variance components of threshold characters by marginal posterior modes and means via Gibbs sampling. *Genet. Sel. Evol.* 27:519–540.
- Hoeschele, I., B. Tier, and H.-U. Graser. 1995. Multiple-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models. *J. Anim. Sci.* 73: 1609–1627.
- Janss, L.L.G., and J. L. Foulley. 1993. Bivariate analysis for one continuous and one threshold dichotomous trait with unequal design matrices and an application to birth weight and calving difficulty. *Livest. Prod. Sci.* 33:183–198.
- Jensen, J., C. S. Wang, D. A. Sorensen, and D. Gianola. 1994. Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. *Acta Agric. Scand.* 44:193–201.
- Moreno, C., D. Sorensen, L. A. Garcia-Cortes, L. Varona, and J. Altarriba. 1997. On biased inferences about variance components in the binary threshold model. *Genet. Sel. Evol.* 29: 145–160.
- SAS. 1989. SAS/STAT® User's Guide. SAS Inst. Inc., Cary, NC.
- Scott, D. W. 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley & Sons, New York.
- Searle, S. R. 1982. Matrix Algebra Useful for Statistics. John Wiley & Sons, New York.
- Sorensen, D. A., S. Anderson, D. Gianola, and I. Korsgaard. 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27:229–249.
- Van Tassell, C. P. 1994. The use of Gibbs sampling for variance component estimation with simulated and weaning weight data using animal and maternal effects models. Ph.D. thesis. Cornell Univ., Ithaca, NY.
- Van Tassell, C. P., G. Casella, and E. J. Pollak. 1995. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. *J. Dairy Sci.* 78: 678–692.
- Van Tassell, C. P., and L. D. Van Vleck. 1996. Multiple-trait Gibbs sampler for animal models: Flexible programs for Bayesian and likelihood-based (co)variance component inference. *J. Anim. Sci.* 74:2586–2597.
- Van Vleck, L. D., and K. E. Gregory. 1996. Genetic trend and environmental effects in a population of cattle selected for twinning. *J. Anim. Sci.* 74:522–528.
- Wang, C. S., R. L. Quaas, and E. J. Pollak. 1997. Bayesian analysis of calving ease scores and birth weights. *Genet. Sel. Evol.* 29: 117–143.



- Wang, C. S., J. J. Rutledge, and D. Gianola. 1993. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet. Sel. Evol.* 25:41–62.
- Wang, C. S., J. J. Rutledge, and D. Gianola. 1994. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet. Sel. Evol.* 26: 91–115.
- Westell, R. A. 1984. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. Ph.D. thesis. Cornell University, Ithaca, NY.
- Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. *J. Dairy Sci.* 71:1310–1318.
- Wright, S. 1934. An analysis of variability in number of digits in an inbred strain of Guinea pigs. *Genetics* 19:506–536.

## Citations

This article has been cited by 5 HighWire-hosted articles:  
<http://jas.fass.org#otherarticles>